Awkward Annie:
Impacts of Playing on the Edge of Social Norms

G. Tanner Jackson

Foundry, Educational Testing Service, Princeton, New Jersey, USA, gtjackson@ets.org

Blair Lehman

Cognitive and Technology Sciences Center, Educational Testing Service, Princeton, New Jersey, USA, blehman@ets.org

Lindsay D. Grace

Interactive Media, University of Miami, Coral Gables, Florida, USA, LGrace@miami.edu

ABSTRACT

Effective interpersonal and cross-cultural communication relies on pragmatics – knowing what to say to whom, and under what circumstances. Nevertheless, pragmatics is generally absent from formal second language instruction. The current effort describes a game designed to assess people's pragmatic ability. In the game, Awkward Annie, players are asked to intentionally select the most inappropriate things to say within conversations (i.e., be inappropriate and see what happens). Thus, players are able to escape from reality by being inappropriate. This work presents a between-subjects study designed to evaluate this twist using two versions of the game (selecting inappropriate versus appropriate responses). Participants in both conditions experienced the same content, but were provided with different goals (be inappropriate, be appropriate). The results indicate that users enjoyed both versions of game equally but performed better within the appropriate version of the game.

CCS CONCEPTS

• Human-centered computing~Empirical studies in HCI • Applied computing~Interactive learning environments

KEYWORDS

Game-based Assessment, Pragmatics, Game Design

1 Introduction

Human communication is a complex balance of meaning and interpretation. Effective communication relies not only on functional word choices and structure, but also on social context (i.e., knowing more than how to put a sentence together, but in knowing what is appropriate to say in a certain situation). Evaluating a person's mastery of language pragmatics is often accomplished through personal interviews (e.g., phone screening during a hiring process). Such interviews are often time consuming, qualitative measures with limited opportunities to assess the wider variety of social contexts with which a person would interact. In short, this work aims to address that challenge, by producing and evaluating a game that addresses this need in pragmatics research.

The value of these assessments is particularly important to an increasingly cross-cultural and interdisciplinary world. The tensions between appropriate and inappropriate language choices become exceedingly relevant as organizations worldwide develop their presence in new communities, employ an increasingly diverse workforce, and seek to engage in dialogue across cultures. It is evident that mastery of language, particularly a second language, is not merely a demonstrated knowledge of the language mechanics, but also a mastery of knowing "when" and "how" to say "what" given specific contexts.

The game Awkward Annie was designed as a first step in creating a reliable means of assessing US English language pragmatics. This paper outlines the educational game research foundation from which the game was designed, the importance of pragmatics assessment, the game's development, and the empirical evaluation of two versions of the game.

1.1 Educational Game Research

Games have been an increasingly popular focus within educational research over the past few decades. Due to their popularity and potential as part of education, the educational game space is extremely large. However, there are still many questions yet to be fully answered or replicated sufficiently to generalize insights across genres. Researchers have been working to synthesize some of the disparate information through meta-analyses that are used to investigate trends across studies, establish generalizable insights, and explore the boundaries and constraints for (un)successful implementations [1,2]. The current effort aims to contribute to the growing body of game research, not through meta-analysis, but through an empirical investigation of a foundational game design choice and exploring the implications of that design on player experience and evidence of educational competency. The empirical focus of this work is to directly address the continued criticism that in such work there is often "more enthusiasm for describing the affordance of games and their motivating properties than for conducting research to demonstrate that these affordances are used to attain instructional aims" [3].

Educational game designers are familiar with the struggle of finding effective ways to align the mechanics of a game with the targeted educational content and skills. Whether mapping via specific frameworks, like MDA [4] or more generally as a concept in practice [5], educational game design has a well-documented demonstration of mapping mechanics to educational content and skills. However, when it comes to heuristics on making such experiences intrinsically interesting and playful, there is little empirical foundation. At this, most research in games relies heavily on the instincts of game designers and the iterative feedback loop of design, implementation, and evaluation. This research aims to aid that cycle, in part, by exploring one of the core game design choices for educational content and assessment. The question is simple – is it more interesting and more playful to design a game that asks players to behave appropriately or inappropriately.

The challenge is also practical, as many games in the education domain aim toward positive practice. Players are tasked with doing the right thing and are rewarded for such. However, in the literature surrounding the psychology of play, researchers heavily emphasize that play is about pushing boundaries, finding the edges of playing against the rules as a means for experimentation [6]. This notion is also evidenced in popular video games, which offer the lure not of stopping at red lights and working a good safe job, but instead of breaking societal rules and in short – behaving badly (e.g., Grand Theft Auto).

The design of Awkward Annie aims to leverage the intrinsically interesting and playful aspects of an educational task while maintaining appropriate educational integrity and maximizing the benefits of a game. By implementing and evaluating both a negative version of the game (i.e., players are rewarded for acting inappropriately) and a positive version of the game (i.e. players are rewarded for acting appropriately) the researchers aim to provide evidence about the positive-negative dichotomy of game design goals.

The game's design and the controlled experiment are intended to examine the effects of this positive-negative design choice in terms of user experience (research question 1) and elicited evidence of pragmatics ability (research question 2).

1.2 Pragmatics and Gaming

In an increasingly globalized world, assessing cross-cultural communication is exceedingly important in the workplace. Historically, the focus of foreign language learning has stressed the fundamentals of a language, such as vocabulary and syntax [7]. However, that instruction often lacks formal training in the more culturally dependent pragmatic aspects of how to say what, to whom, when, and where [8]. Pragmatics is one of the fields that directly relates to a person's ability to communicate effectively and appropriately within a given language and cultural context [9]. Research has illustrated the importance of success and failures in pragmatics within everyday interactions [10,11,12]. Pragmatics, as an academic field, relates to contextualized social interaction variables and therefore representation of these interactions requires rich settings and nuanced inputs (potentially well suited to games). Unfortunately, the field of pragmatics lags behind others in terms of educational games [13] with an exception being Julie Sykes' work on Spanish pragmatics [14].

There are many aspects to pragmatics and how to successfully interact in a particular social and cultural context. In particular, there are two primary aspects of pragmatics: pragmalinguistics and sociopragmatics. Pragmalinguistics relates to the specific words and phrases used to communicate (how to say what), whereas sociopragmatics relates more to social situations and context (what, who, and where). These skills are typically assessed using interviews, which are both labor intensive and challenged by scalability and consistency. Games could potentially address these three limitations of existing assessments, but very few researchers have investigated how games could be leveraged to improve assessment in this area (with [14] being an exception). The benefits of such games include their scalability (they can be deployed to large populations), their engagement potential (they are considered more engaging than traditional education) and their consistency [15]. The current work was motivated by this combination of factors that include a skill that has impact on everyday interactions, lack of formal training materials, dearth of education games in this space, and potential alignment with game mechanics, which eventually led to the design of Awkward Annie.

2 Awkward Annie

Given the complex and contextualized nature of pragmatics, the design team explored many potential game designs that incorporated the rich contextual variables (separately as minigames and together as a single game), faithfully represented the constructs (i.e., face validity), provided evidence of a player's ability (interpretable success/failure), and was enjoyable to play. Through an iterative design process the team decided to pursue a single game that could represent multiple types of relationships, different levels of status, and different kinds of interactions: Awkward Annie.

Awkward Annie is an educational game focusing on English language pragmatic skills and represents a novel design that asks players to violate social norms by intentionally selecting the most inappropriate things to say to virtual colleagues (see screenshot in Figure 1). The intentionally awkward social experience involves players controlling [Awkward] Annie, who is a weird but endearing character that fails at social interactions. Annie continually puts her friends and colleagues in awkward situations by saying the weirdest and most inappropriate things (selected by the player). This design allows players to apply their English language pragmatics skills (sociopragmatics in particular) in a playful way as they engage in numerous conversations with characters representing old friends, first time acquaintances, same-level colleagues, and new bosses. The domain of pragmatics focuses on social interactions and how to adapt to particular situations. Thus, the game was designed to focus on multiple conversational interactions that embody various social practices common in everyday interactions (greetings, small talk, making requests) and have important dimensions

that modify when some things are more or less appropriate (familiarity with someone, status differences between colleagues, and the size/sensitivity of a particular request or topic).

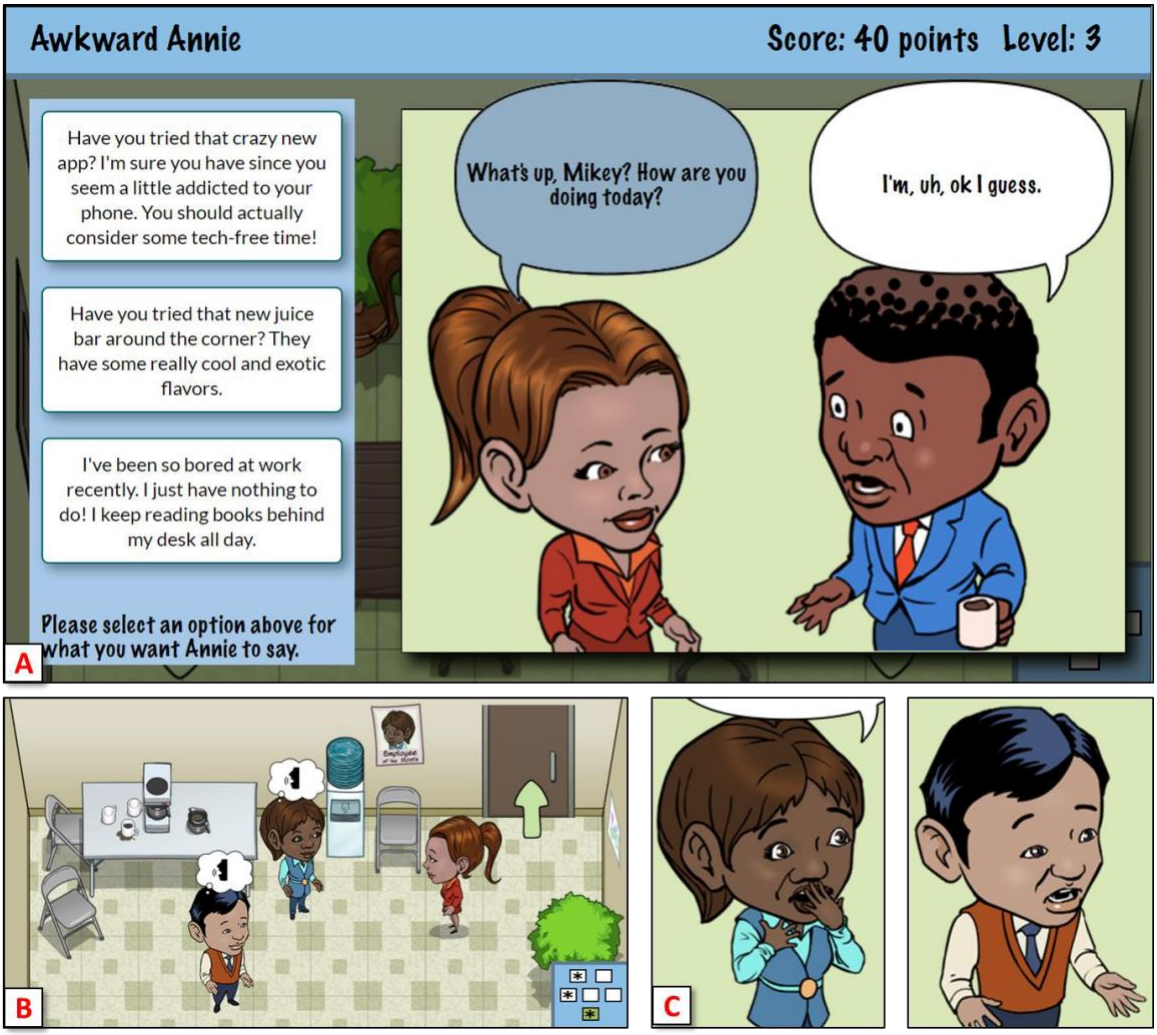

Figure 1: Screenshot examples of Awkward Annie (A) conversations, (B) navigation, and (C) character reactions

Within the game, users navigate around and approach one of four characters to initiate new conversations. As conversations are completed successfully the player moves up in levels, which open up new conversations with the characters; however, the conversations were not designed to increase in difficulty across levels in this version of the game (i.e., difficulty was designed to remain constant across levels). Each conversation lasts three turns (i.e., the player makes three decisions) and each turn involves selecting one of three conversational options that Annie will say to the other character. In the original version of the game, players are asked to select the most inappropriate response option (i.e., confirm the negative evidence) and thus indicate their knowledge by not selecting the appropriate option. For example, in Figure 1A the top option is considered correct and is the most inappropriate (impolite + prescriptive statement), with the second option being appropriate (polite + relevant), and the bottom option being the moderately inappropriate option (impolite or over-sharing information).

Each selection is associated with a specific point value and those points are used both as a scoring mechanism for the game overall and for determining success within a given conversation. A correct selection will earn the player five points, a suboptimal selection (i.e., moderately inappropriate) earns three points, and an

incorrect selection earns zero points. Prior norming studies were conducted with a native US English speaking population to evaluate the conversation options and content was refined to ensure adequate "distance" to allow for discrimination between the three different response options in each conversation. Thus, a given score for a conversation indicates a player's ability to discriminate between optimal and suboptimal responses. Players must accumulate ten or more points to successfully complete each conversation (maximum of 15 points for a set of perfect selections). A player can choose from any available conversations within the environment, but failed conversations must be retried until successful in order to progress to the next conversation for that character. All conversations for all characters (22 total across characters) must be successfully completed to finish the game. Thus, the proportion of successful conversations (# of successful conversations / # of total conversations) is an indicator of general progress and ability (with higher proportions indicating higher ability through fewer failures).

2.1 Previous Studies with Awkward Annie

The current study is the fourth effort in a series of studies with Awkward Annie [16,17]. The first study explored the potential efficacy of Awkward Annie with non-native English speakers (i.e., is there enough potential to continue the work) while the second study involved native US English speakers to establish a performance benchmark. Based on player feedback in both studies, a third study was conducted that focused on alignment and norming of content. This third study was conducted with 200 native US English speakers who piloted a non-game pragmatics measure (to be used in the current study) and reviewed the game content (including appropriateness of the artwork) to rate conversation options on their perceived (in)appropriateness. Based on the native speaker ratings and input from a subject matter expert (Timpe-Laughlin), the pragmatics measure was adjusted to improve item functioning (example in Figure 2) and the game content was modified to ensure that conversation options were developed in a consistent manner so that native US English speakers could adequately discriminate between options (i.e., if native speakers cannot discriminate between options then non-native speakers couldn't be expected to either). The current (fourth) effort was designed to test a fundamental design tension identified during development (i.e., being inappropriate is more fun than being appropriate, but must provide evidence of player abilities).

3 Study

The current study was designed to explore one of the fundamental game design choices of Awkward Annie (inappropriate response selection) and how that choice may impact both user experience and the assessment evidence collected through in-game performance. Specifically, the current study manipulated the goal provided to the player, which is to either select the most appropriate conversational response (confirming the "positive" evidence) or select the most inappropriate response (confirming the "negative" evidence). The original game design decision was to have users play on the edge of social norms and escape from normal day-to-day interactions by selecting the most inappropriate response. Thus, our first research question (RQ1) is, how does the conversational goal (appropriate vs. inappropriate) impact the user experience within Awkward Annie?

Similarly, the unconventional nature of selecting the most inappropriate response also has implications for how well the user's pragmatic ability can be estimated. Selecting the inappropriate response requires a scoring inference that the user knows the appropriate response by not selecting it. Additionally, selecting which response is the most inappropriate (out of 1 appropriate and 2 inappropriate) may be more nuanced and difficult than just selecting the one correct response. Thus, our second research question (RQ2) is, how does the conversation goal (appropriate vs. inappropriate) affect the player performance evidence collected within Awkward Annie?

3.1 Materials

The current study focused on examining users' pragmatics skills and experience with Awkward Annie. Awkward Annie was used as one measure of users' pragmatic skills. The other measure of pragmatic skills was a selected-response test (traditional test) consisting of short stimuli and multiple-choice answer options (see sample item in Figure 2). Specifically, the traditional test presented users with a short two or three sentence scenario that included contextual information relating to setting, familiarity between speakers, and the intent of the type of interaction (e.g., greeting, engaging in small talk, or making a request). After reading the scenario and example conversation, users were asked to rate a particular response on a five-point scale from "Very Impolite (Very Inappropriate)" to "Far Too Polite (Very Inappropriate)" with a midpoint of "Polite (Appropriate)". Test takers' responses were then compared to the distribution of native speaker responses.

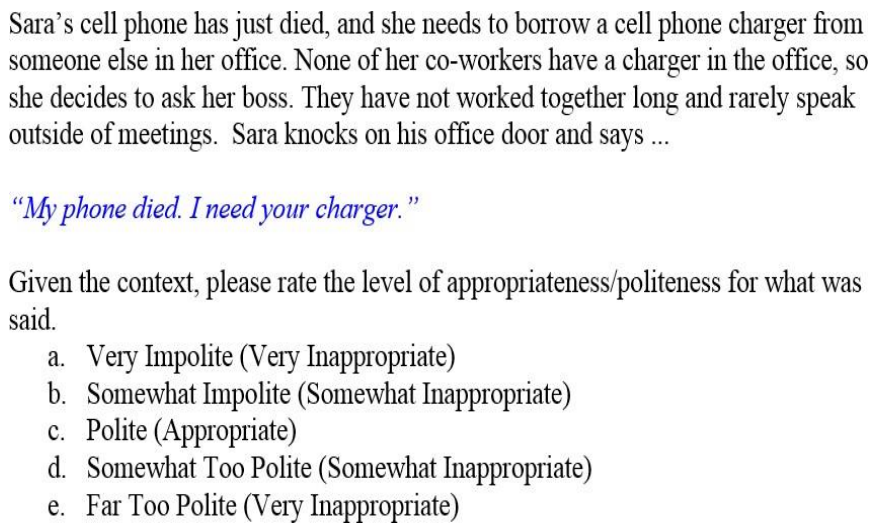

Figure 2: Example item from the traditional test of pragmatics.

For the current study, user experience was broadly defined as a set of affective variables and environment perceptions. To examine users' experience, a survey was administered immediately prior to (pre) and after (post) playing Awkward Annie. The set of user experience questions included twelve items on a 10-point scale (with higher values indicating a stronger feeling) that addressed user emotions and perceptions. Players rated how much they felt amused, anxious, bored, confused, curious, embarrassed, engaged, enjoyment, and frustrated. These emotions were selected because they have been found to be important in non-educational games [18,19,20,21], educational games [22,23,24], and assessments [25]. Additionally, players rated their perceptions of the environment as being more work-like or play-like, how likely it was that they would play the game again, and if they liked the game overall.

3.2 Sample

The participant sample for the study consisted of 350 native US English speaking adults who currently live in the US and were recruited via Amazon Mechanical Turk. Over half (56%) of participants were male (43% female, <1% other or not respond), the majority (49%) were between 31 and 45 years old (under 21yrs: <1%, 21-30yrs: 37%, over 46yrs: 14%), and almost all grew up in the US (US: 99%, other: 1%).

3.3 Sample

The online participants each received $15 after completing the three study phases: pretest, gameplay, and posttest. The pretest was designed to collect basic demographic information, English language pragmatics

skills (traditional test), and a baseline (pretest) for user experience variables (to allow for comparison at posttest).

For the gameplay phase participants were randomly assigned (between subjects) to either the appropriate or inappropriate version of Awkward Annie. Both versions used the same conversation content and differed only in terms of the goal provided to the user (appropriate vs. inappropriate). Both versions of the game included 22 conversations across the characters (5 conversations per character and 2 introductory conversations that provided background). On average, games lasted slightly more than 20 minutes for both conditions.

Upon completing the game, participants transitioned to the posttest where they answered questions about their experience within the game (user experience questions matched with those from the pretest). Participants completed the full study in one hour-long session.

3.4 Results

After completing the online data collection, a series of analyses was conducted to investigate issues related to the Awkward Annie system and to address our primary research questions (RQ1 and RQ2 above). Initial analyses were conducted to determine if there were any differences between the two conditions prior to playing Awkward Annie. These initial analyses included comparing pretest user experience and performance on the traditional test. Then, our analyses focused on the posttest user experience (RQ1) and in-game evidence of performance (RQ2).

The comparison of pretest variables allows for the determination of whether or not the two condition groups can be treated as equivalent. An independent samples *t*-test revealed that the appropriate (M=.80, SD=.07, n=177) and inappropriate conditions (M=.79, SD=.08, n=173) did not differ on pragmatics skills as measured by the traditional test, $t(348)$=-.78, p=.44. Similarly, the two conditions were equivalent on pretest user experience (*p's*>.05), with one exception. The appropriate condition (M=7.62, SD=2.15) reported significantly higher levels of pretest engagement than the inappropriate condition (M=7.11, SD=2.54), $t(348)$=-2.01, p=.04 (though with a fairly small effect size, d=.22). Thus, the two conditions seem to be roughly equivalent on levels of pragmatic skills and user experience immediately prior to their interaction with Awkward Annie.

Next, the first research question was investigated. After the game, participants in both conditions rated both game versions similarly (*p's*>.05), with higher ratings for amused, engaged, enjoyment, and play-like, lower ratings for bored, and near zero ratings for anxious, confused, embarrassed, and frustrated (see dark bars in Figure 3). Thus, there appear to be no differences between the users' experiences for the appropriate versus inappropriate version of the game (RQ1).

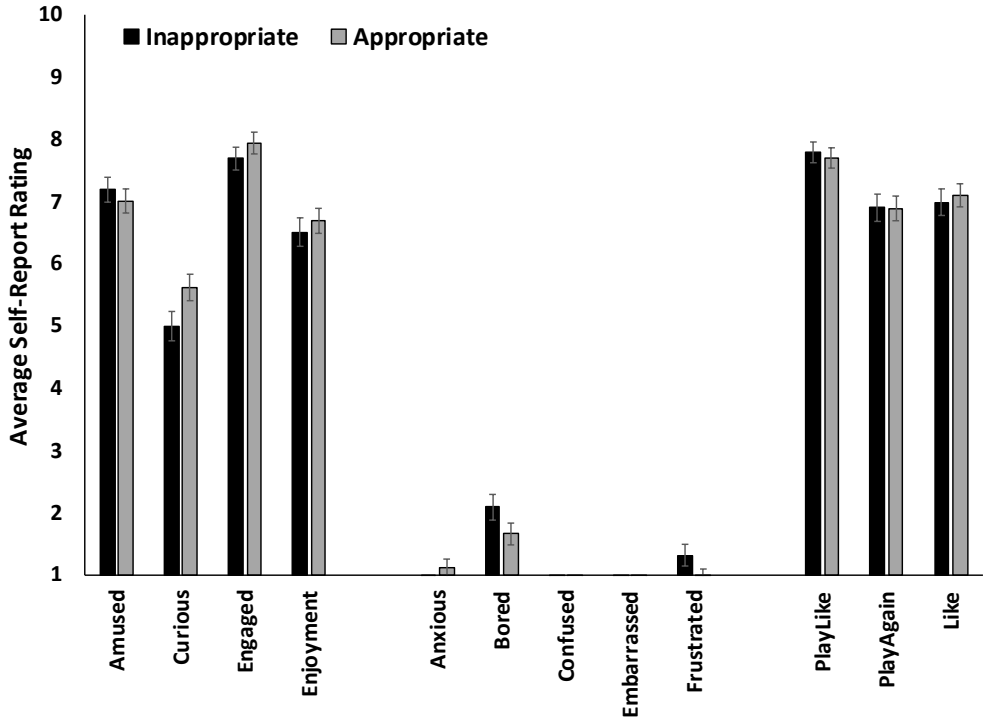

Figure 3: Comparison of user experience across conditions.

In addition, to investigating the difference in user experience between the two game conditions, this work also explored the difference in user experience from before (pretest) and after (posttest) playing Awkward Annie. This comparison allows for us to determine whether or not there was an overall positive experience for completing Awkward Annie, and given that the pretest user experience questions were administered just after the traditional test, the difference from pretest to posttest allows for an initial comparison of the traditional approach to the game approach (though a counterbalanced design would be needed for stronger conclusions in that regard, though that is not the focus of this work). A mixed-design ANOVA that compared user experience variables between conditions from pretest to posttest revealed no significant differences between conditions, significant differences between almost all user experience variables over time (with the exception of confusion and anxiety which were stable), and no interaction between condition and time. Figure 4 shows a comparison of the pretest and posttest user experience variables. The main effect from pretest to posttest reveals that regardless of game condition, participants had a positive experience completing Awkward Annie (and likely much more so than completing the traditional test – though a counterbalanced design would be needed for a true comparison).

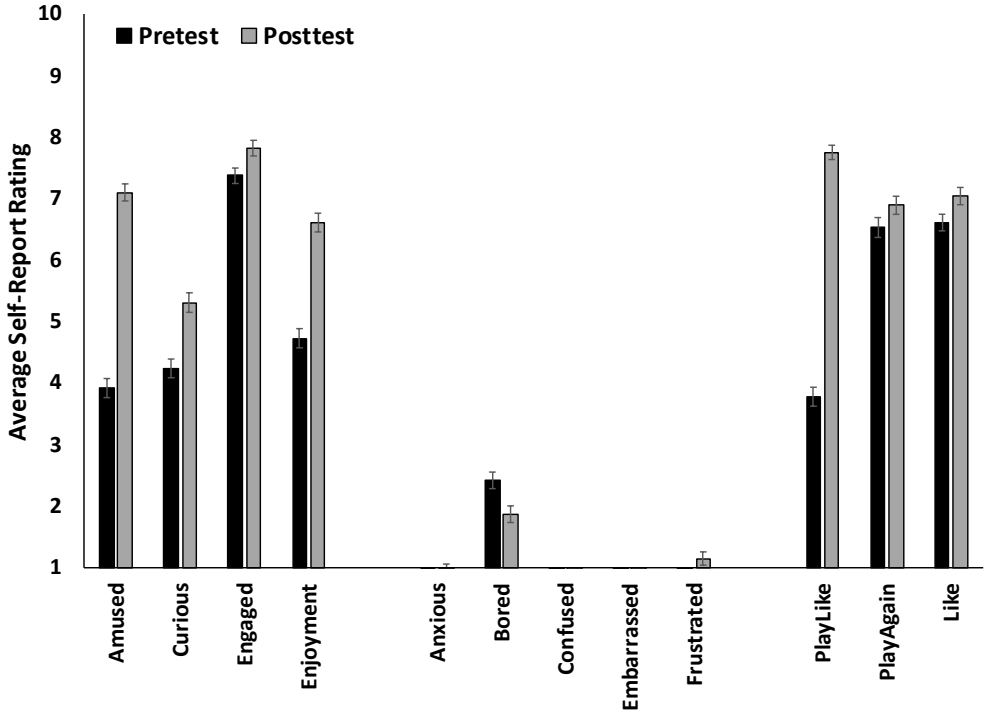

Figure 4: User experience overall on the pre- and posttest.

Beyond user experience, analyses were also conducted to explore in-game performance across game conditions and how that performance may relate to other variables (e.g., pragmatics skills, user experience). Two metrics were used as primary indicators of game performance: proportion of correct conversations and average conversation score. The proportion of successful conversations was calculated separately for each user (# successful conversations / total # of conversations attempted) and represents overall ability at a coarse-grained level. Average conversation scores were also calculated separately for each player and included the scores from all conversations (including both successful and failed conversations). The average conversation score represents a player's ability to discriminate between optimal and suboptimal response options (with a range of 0 to 15 points and higher scores indicating more accurate discrimination within the game).

Analyses were used to compare in-game performance between the two game versions and found that participants in the appropriate condition of the game scored significantly higher on both performance metrics than those in the inappropriate condition (see Table 1). Given no differences between the groups on the pretest of pragmatic skills, the performance differences within the game seem to indicate that the appropriate version was easier than the inappropriate version. Thus, there does appear to be some potential difference in terms of what information the versions of the games elicit from players (RQ2). This high rate of success is expected for native speakers because pragmatic issues often receive feedback during initial language development (e.g., parents correcting their children on how to phrase a request appropriately), whereas pragmatics feedback and instruction does not often occur during formal second language learning opportunities [8]. Prior work with the inappropriate version of Awkward Annie has found that, indeed, non-native English speakers do score significantly lower than native English speakers on both game performance metrics [16].

Table 1. Game performance means (SDs) across conditions.

	Proportion of Successful Conversations	Average Conversation Score
Appropriate Version	.98 (.04)	14.23 (.80)
Inappropriate Version	.94 (.08)	12.65 (1.24)
$t_{version}$	6.36	14.27
p	<.001	<.001

Further analyses were conducted to examine the relation between the two game performance measures and the traditional test scores. Pearson correlations revealed that the traditional test scores were significantly correlated with both the proportion of successful conversations for the appropriate (r=.302, p<.00) and inappropriate version (r=.278, p<.001) as well as the average conversation scores for the appropriate (r=.382, p<.001) and inappropriate version (r=.313, p<.001). Thus, the game does seem to be measuring some similar aspects of pragmatics as the traditional test.

4 Discussion

The current work was designed to explore a fundamental design decision for an educational game through user experience and game performance. A sample of 350 native English-speaking adults played one of two versions of Awkward Annie (appropriate or inappropriate). Players did not report any user experience differences between games (RQ1) but performed significantly better in the appropriate version than the inappropriate version (RQ2).

Analyses on the user experience ratings revealed that both versions of the game were reported to be amusing, engaging, and enjoyable without being boring, confusing, or frustrating. It is clear that both games are viewed positively, but somewhat surprising that there were no differences between the conditions on any of the user experience dimensions. There are at least two possible explanations for this finding. First, because the game uses the same content, players in both versions see the same content options and are exposed to the same humorous situations and response options. Thus, the humor alone could be carrying the effect across both conditions. Second, the experiment was conducted online through Amazon Mechanical Turk and this particular task is likely much more enjoyable than the typical online tasks like surveys and non-interactive activities that participants are asked to complete (in fact participants frequently commented and provided feedback to this effect). Thus, the act of selecting the most inappropriate response option may be less critical of a design factor than simply providing the humorous content and allowing the players to imagine it as an option to consider.

Players in both the appropriate and inappropriate versions of the game performed very well with 98% and 94% of conversations being successful, respectively. Interestingly, players in the appropriate version had an average conversation score of 14.23 (95% out of a perfect 15), whereas players in the inappropriate version had a lower average conversation score of only 12.65 (84% out of a perfect 15). Based on the performance between conditions in this study, it appears that for highly proficient (i.e., native) speakers, the inappropriate version of the game is slightly more difficult than the appropriate version. This has interesting implications for non-native speakers and suggests that the different versions of the game could be more (or less) appropriate for varying levels of English proficiency. Along these lines, previous Awkward Annie studies with the inappropriate version found that non-native English speakers performed with an average conversation score of only 11.82 (78% out of a perfect 15; [16]) in the inappropriate version. This trend tentatively supports the idea that the game could make use of the same content but varying the goal (appropriate/inappropriate) could provide a more suitable match based on the English proficiency of non-native speakers (e.g., the appropriate version may be more suitable for lower proficiency speakers).

Limitations of the current work include a lack of counterbalancing for the external pragmatics test, which would have facilitated a more direct comparison between a traditional test to the two game-based approaches for assessing pragmatics. Future work will more thoroughly investigate and compare these two approaches (i.e., game and non-game assessments of pragmatics). Additionally, the current work leveraged a sample of native English speakers due to the assumption that the fundamental design decision to include inappropriate content would be a fairly generalizable effect across populations. Given the performance differences between conditions and the lack of user experience differences within the native speaking population subsequent efforts will explore effects with various non-native English-speaking populations.

Despite these limitations, the user experience results here were surprising in the fact that the players did not have to enact the inappropriate choices in order to receive the same positive benefits. It appears that merely being exposed to the humorous options was sufficient to positively impact affect, and thus engaging in that socially taboo behavior did not contribute significantly beyond exposure (at least in this case). This is likely, at least in part, due to the fact that this was a between-subjects study in which users did not know what they were missing and had no basis for comparison other than the traditional test at the beginning.

Overall, the work reported here explored a foundational game design decision of Awkward Annie and has implications both for game designers (relative impacts of content versus action) and educational researchers (performance differences based on evidence identification requirements that reuse content).

ACKNOWLEDGMENTS

The researchers would like to acknowledge John S. and James L. Knight Foundation for their financial support in publishing this research.

REFERENCES

[1] Douglas B. Clark, Emily E. Tanner-Smith, and Stephen S. Killingsworth. 2016. Digital games, design, and learning: A systematic review and meta-analysis. *Review of Educational Research* 86, 1 (March 2016), 79-122. https://doi.org/10.3102/0034654315582065

[2] Michael F. Young, Stephen Slota, Andrew B. Cutter, Gerard Jalette, Greg Mullin, Benedict Lai., . . . Mariya Yukhymenko, 2012. Our princess is in another castle: A review of trends in serious gaming for education. *Review of Educational Research* 82, (March 2012), 61– 89. DOI:10.3102/0034654312436980

[3] Sigmund Tobias, J. D. Fletcher, David Yun Dai, and Alexander P. Wind (2011). Review of research on computer games. In *Computer Games and Instruction*. Information Age Publishing, Charlotte, NC, 127–222.

[4] Robin Hunicke, Marc LeBlanc, and Robert Zubek. "MDA:A formal approach to game design and game research." *Proceedings of the AAAI Workshop on Challenges in Game AI*. Vol. 4. No. 1. 2004.

[5] James Paul Gee (2011). "Reflections on empirical evidence on games and learning." *Computer games and instruction* (2011): 223-232.

[6] Robert Caper (1996). Play, experimentation and creativity. *The International journal of psycho-analysis*, 77.5: 859.

[7] Frederik De Grove, Peter Mechant, and Jan Van Looy. 2010. Uncharted waters?: Exploring experts' opinions on the opportunities and limitations of serious games for foreign language learning. In *Proceedings of the 3rd ACM International Conference on Fun and Games*. ACM Press, New York, NY, 107-115.

[8] Noriko Ishihara and Andrew D. Cohen, 2010, *Teaching and Learning Pragmatics: Where Language and Culture Meet* (1st ed.). Pearson, New York, NY.

[9] Kathleen Bardovi-Harlig. 2013. Developing L2 pragmatics. *Language Learning* 63, 1 (March 2013), 68-86. DOI: 10.1111/j.1467-9922.2012.00738.x

[10] N. Riddiford and A. Joe (2010). Tracking the development of sociopragmatic skills. *TESOL Quarterly*, 44(1), 195-205.

[11] M. Clyne. 1994. Inter-cultural communication at work: Cultural values in discourse. Cambridge University Press, Cambridge, UK.
[12] Veronika Timpe. 2013a. *Assessing intercultural language learning*. Frankfurt a.M.: Peter Lang.
[13] N. Taguchi and C. Roever. 2017. *Second language pragmatics*. Oxford: Oxford University Press.
[14] J. Sykes, & J. Reinhard. (2012). *Language at Play: Digital Games in Second and Foreign Language Teaching and Learning*. Series on Theory and Practice in Second Language Classroom Instruction, J. Liskin-Gasparro & M. Lacorte, series eds. Pearson-Prentice Hall.
[15] James Paul Gee. 2003. What video games have to teach us about learning and literacy. *Computers in Entertainment (CIE)-Theoretical and Practical Computer Applications in Entertainment* 1, 1 (Oct. 2003), 1-4. DOI: 10.1145/950566.950595
[16] Jackson, G. T., Grace, L., Inglese, P., Wain, J., & Hone, R. (2017, December). Awkward Annie: Game-Based Assessment of English Pragmatic Skills. In *International Conference on Advances in Computer Entertainment* (pp. 795-808). Springer, Cham.
[17] Grace, L., Jackson, G. T., & Lehman, B. (2020, April). The Effect of Positive and Negative Goal Orientation and Player Embodiment in Assessment Games. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems* (pp. 1-8).
[18] K. Gilleade and A. Dix. (2004). Using frustration in the design of adaptive videogames. In *Proceedings of the 2004 ACM SIGCHI International Conference on Advances in computer entertainment technology* (pp. 228-232). New York, NY: Association for Computing Machinery.
[19] N. Lazzaro. (2004). *Why we play games: four keys to more emotion without story*. Paper presented at Game Developers Conference. San Jose, California.
[20] J. Lumsden, E. Edwards, N. Lawrence, D. Coyle, and M. Munafò. (2016). Gamification of cognitive assessment and cognitive training: A systematic review of applications and efficacy. *JMIR Serious Games, 4*, e11.
[21] A. Nylund and O. Landfors. (2015). Frustration and its effect on immersion in games: A developer viewpoint on the good and bad aspects of frustration. (Unpublished master's thesis). Umeå University, Sweden.
[22] S. McQuiggan, J. Robison and J. Lester. (2010). Affective transitions in narrative-centered learning environments. *Educational Technology & Society, 13*, 40-53.
[23] M. Rodrigo and R. Baker. (2011). Comparing the incidence and persistence of learners' affect during interactions with different educational software packages. In R. Calvo and S. D'Mello, (Eds.), *New perspectives on affect and learning technologies* (pp. 183-202). New York, NY: Springer.
[24] M. Rodrigo and R. Baker. (2014). Comparing learners' affect while using an intelligent tutor and an educational game. *Research and Practice in Technology Enhanced Learning, 6*, 43-66.
[25] Blair Lehman and Diego Zapata-Rivera. (2018). Student emotions in conversation-based assessments. *IEEE Transactions on Learning Technologies, 11*, 1-13.
[26] Patricia S. Abril and Robert Plant. 2007. The patent holder's dilemma: Buy, sell, or troll? *Commun. ACM* 50, 1 (Jan. 2007), 36-44. DOI: https://doi.org/10.1145/1188913.1188915